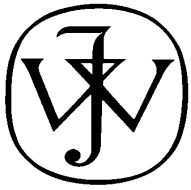

WILEY

Publishers Since 1807

111 River Street, Hoboken, NJ 07030

*****IMMEDIATE RESPONSE REQUIRED*****

Please follow these instructions to avoid delay of publication.

☐ **READ PROOFS CAREFULLY**

- This will be your only chance to review these proofs.
- Please note that the volume and page numbers shown on the proofs are for position only.
- If you have submitted your manuscript on computer disk, you still need to proofread the page proofs carefully against the original manuscript.

☐ **ANSWER ALL QUERIES ON PROOFS** (Queries for you to answer are noted on the manuscript.)

- Mark all corrections directly on the proofs, not on the manuscript. Note that excessive author alterations may ultimately result in delay of publication and extra costs may be charged to you.

☐ **CHECK FIGURES AND TABLES CAREFULLY** (Color figures will be sent under separate cover.)

- Check size, numbering, and orientation of figures. Check quality of figures directly from the galley proofs. The reproduction is 1200dpi and indicative of final printed quality.
- Review figure legends to ensure that they are complete.
- Check all tables. Review layout, title, and footnotes.

☐ **COMPLETE REPRINT ORDER FORM**

- Fill out the attached reprint order form. It is important to return the form even if you are not ordering reprints. You may, if you wish, pay for the reprints with a credit card. Reprints will be mailed only after your article appears in print. This is the most opportune time to order reprints. If you wait until after your article comes off press, the reprints will be considerably more expensive.

RETURN

- ☐ **PROOFS**
- ☐ **ORIGINAL MANUSCRIPT**
- ☐ **REPRINT ORDER FORM**
- ☐ **ORIGINAL FIGURES (If included)**
- ☐ **CTA (If you have not already signed one)**

Express Mail complete package to:

Margaret Ziomkowski
Professional/Trade Group
c/o John Wiley & Sons
111 River Street, 5-01
Hoboken, NJ 07030-5774
Fax: 201-748-8824
E-mail: MZiomkow@wiley.com

RETURN WITHIN 48 HOURS OF RECEIPT VIA EXPRESS MAIL
AUTHORS OUTSIDE THE U.S. YOUR MANUSCRIPT AND PROOFS HAVE BEEN SENT TO YOU VIA THE QUICKEST MANNER POSSIBLE. PLEASE ASSIST US IN KEEPING YOUR ARTICLE ON SCHEDULE BY RETURNING THE MANUSCRIPT, PROOFS, AND FIGURES TO US AS QUICKLY AS POSSIBLE. YOUR COOPERATION IS GREATLY APPRECIATED.

☐ **IF THIS BOX IS CHECKED, YOUR ARTICLE IS BEING TREATED AS A RUSH AND NEEDS TO BE RETURNED WITHIN 24 HOURS**

111 River Street, 5-01
Hoboken, NJ, 07030-5774
TEL (201) 748-6599
FAX (201) 748-8824

Notice to Authors

Enclosed please find the proofs for your article. No further proofs will be sent to you.

Please return the corrected proofs within 48 hours of receipt to ensure prompt publication. If your proofs are not returned within this period we reserve the right to publish your paper as typeset with no further corrections. The author is solely responsible for any errors that are undetected by his/her proofreading.

You alone are responsible for checking the proofs of your article and for making all necessary corrections on the proofs. Please be sure to answer all queries posed in the margins of the proof. ***We reserve the right to charge authors*** for any alterations to the text or illustrations which are not due to typesetting errors, at a rate of no less than $2.00 per affected line. No further changes may be made once the marked proofs have been returned to us for processing.

Unless otherwise indicated, return the enclosed material (marked proofs) within 48 hours by express mail, fax, or e-mail to: Margaret Ziomkowski (MZiomkow@wiley.com).

Professional/Trade Group
Wiley Periodicals
C/O John Wiley & Sons
111 River Street, 5-01
Hoboken, NJ 07030-5774
U.S.A.

To order reprints of your article, complete the order form enclosed. The reprint order form MUST be returned with your proofs. Payment may be made by personal check, credit card, or a purchase order from your institution. Make sure a copy of this form is returned with your purchase order or payment. International orders must be paid in U.S. currency and drawn on a U.S. bank. Reprints should be ordered prior to publication since post publication orders are subject to a back-to-press surcharge.

Proofreader's Marks

Mark in Margin	Mark in Proof	Meaning
cap /	are capitalized in english titles	Capital letter
lc /	only the first word and Proper	Lower-case letter
tr /	nouns in French and Italian	transpose
^ /	titles and all nouns in German.	Insert comma
⁀ /	Setting the vollume number of	Delete and close up
X /	a periodical in distinctive type	Fix broken letter
⁀ /	(italics or bold face) and page	Delete hyphen; leave space
" /	"Vol and "p" unnecessary	Insert quotes
= /	The year of publication is en=	Insert hyphen
e /	closed in parenthesis only when	Substitute e for i
eq # /	information as / when, / for ex-	Equalize spacing
[l. l /	[ample, the volume and page of	Move left
stet /	a periodical ~~are~~ cited and the	Let type stand
w.f.	year is added for the convenience	Wrong type font
¶ /	of the reader. ¶ The following	Begin new paragraph
:/	illustrates a bibliographic item/	Change period to colon
S.C. /	Gay, Charles M, and Charles	Set in small caps
ital /	De van Fawcett, *Mechanical* and	Set in italics
for /	*Electrical Equipment of Buildings*	Change of to for
rom /	John Wiley & Sons, New York	Set in roman
C / # /	second edition, p 229, 1945	Close up, insert space

WILEY
Publishers Since 1807

REPRINT BILLING DEPARTMENT • 111 RIVER STREET • HOBOKEN, NJ 07030
PHONE: (201) 748-8789; FAX: (201) 748-6326
INTERNET: reprints@wiley.com

PREPUBLICATION REPRINT ORDER FORM

To order reprints, please complete this form. If you do not want reprints, please fill in your name and your article's title, cross out the rest of the form, and return it with your proofs. Reprints are shipped 4 to 6 weeks after publication. Reprints ordered after printing are much more expensive.

JOURNAL: ***JOURNAL OF CLINICAL PSYCHOLOGY / IN SESSION*** VOLUME ____ ISSUE ____
TITLE OF MANUSCRIPT _____
MS. NO. _____ NO. OF PAGES _____ AUTHOR (S) _____

REPRINTS 7 X 10

No. of Pages	100 Reprints	200 Reprints	300 Reprints	400 Reprints	500 Reprints
	$	$	$	$	$
1-4	193	354	496	627	747
5-8	246	454	642	814	959
9-12	303	557	782	996	1,178
13-16	385	660	927	1,177	1,392
17-20	468	761	1,073	1,353	1,604
21-24	524	864	1,212	1,540	1,815
25-28	591	967	1,357	1,722	2,035
29-32	672	1,068	1,498	1,910	2,254
33-36	739	1,172	1,647	2,091	2,467
37-40	798	1,274	1,788	2,273	2,679

** REPRINTS ARE ONLY AVAILABLE IN LOTS OF 100. IF YOU WISH TO ORDER MORE THAN 500 REPRINTS, PLEASE CONTACT OUR REPRINTS DEPARTMENT AT (201) 748-8789 FOR A PRICE QUOTE.

COVERS

100 Covers - $90 • 200 Covers - $145 • 300 Covers - $200
400 Covers - $255 • 500 Covers - $325 • Additional 100s - $65

☐ Please send me _____ reprints of the above article at …… $ _____
☐ Please send me _____ covers of the above article at …….. $ _____
Please add appropriate State and Local Tax {Tax Exempt No._____} $ _____
Please add 5% Postage and Handling ……. $ _____
TOTAL AMOUNT OF ORDER ** …… $ _____
*** International orders must be paid in U.S. currency and drawn on a U.S. bank*

Please check one: ☐ Check Enclosed ☐ Bill Me ☐ Credit Card
If credit card order, charge to: ☐ American Express ☐ Visa ☐ MasterCard ☐ Discover
Credit Card No. _____ Signature _____ Exp. Date _____

Bill To:
Name _____
Address _____
_____
_____
_____
Purchase Order No._____

Ship To:
Name _____
Address _____
_____
_____
_____
Phone _____ Fax _____
Internet _____

YOU CAN HELP US ACHIEVE THE LARGEST POSSIBLE READERSHIP FOR YOUR ARTICLE

Please take a moment to give us the names and addresses of any organizations or companies that might be interested in purchasing reprints of your article, and return this form along with your page proofs. If you prefer to send it separately, please mail it to:

Wiley Periodicals, Inc.
c/o John Wiley & Sons
111 River Street
Hoboken, NJ 07030
Attn: Reprints Department

Your Article Number _____

Organization _____

Address_____

_____

_____

_____

Contact Name _____

Phone _____

Fax _____

Organization _____

Address_____

_____

_____

_____

Contact Name _____

Phone _____

Fax _____

QUERIES

1. Correct idea?

2. Please update.

3. Correct year for survey?

4. Please update.

5. Lowercase OK or is this a proprietary name requiring capitalization?

6. Change from or correct? Two methods meant?

7. Change from proceeding correct? Or is preceding meant?

8. Please update.

A Lab Test and Algorithms for Identifying Clients at Risk for Treatment Failure

Corinne Hannan, Michael J. Lambert, Cory Harmon, Stevan Lars Nielsen, David W. Smart, and Kenichi Shimokawa
Brigham Young University

Scott W. Sutton
Emory and Henry College

Many branches of medicine rely heavily on lab tests to monitor client treatment response and use this information to modify their treatment. By contrast, those who offer psychological interventions seldom rely on formal assessments (lab tests) to monitor their clients' response to treatment. Data are presented that demonstrate that clinicians rarely accurately predict who will not benefit from psychotherapy. This finding is contrasted with the use of a questionnaire (lab test data) and decision rules on the basis of a client's expected progress. Results have indicated that formal methods of monitoring were able to identify 100% of the patients whose condition had deteriorated at termination, and 85% by the time they had attended three treatment sessions. Practitioners are encouraged to consider formal methods of identifying the deteriorating client. © 2005 Wiley Periodicals, Inc. J Clin Psychol/In Session 61: 1–9, 2005.

Keywords: deterioration; negative outcomes; Outcome Questionnaire-45; identification of treatment failure; outcomes management; psychotherapy outcome

Lab tests used in conjunction with empirically based cut scores and decision trees are regularly employed in various branches of medicine. These tests are used particularly when clinical judgment is involved in the interpretation of diverse information because lab test results are generally more accurate than more subjective methods. In the diagnosis and treatment of prostate cancer, for example, a physician commonly begins by taking

Correspondence concerning this article should be addressed to: Michael J. Lambert, Department of Psychology, 272 TLRB, Brigham Young University, Provo, UT 84602; e-mail: mike_lambert@byu.edu

JCLP/In Session, Vol. 61(2), 1–9 (2005)

Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jclp.20108

a blood sample and ordering a prostate-specific antigen (PSA) test as a chemical marker for the probability of cancer. Since introduction of this lab test in 1987 an enormous increase in the reported incidence of prostate cancer has been observed. Scores on the PSA test higher than 4.0 nanograms per milliliter raise concern, prompt further testing, and can trigger more invasive and expensive testing, such as a prostate biopsy. Administration of each test increases the likelihood of making an accurate diagnosis and maximizing the success of treatment. But even after treatment (e.g., chemotherapy, radiation, surgery) is undertaken, monitoring of PSA continues in order to provide a marker for treatment success or failure and to guide the need for further action. In contrast, there are no agreed upon lab tests commonly used by mental health practitioners that serve as a signal for success and failure during psychotherapy. Such a lab test could supplement, but not replace, clinical decision making during the course of psychotherapy.

In ordinary circumstances clinicians informally monitor treatment progress, making adjustments in their behavior in accordance with the characteristics of their clients, theoretical considerations, as well as their impressions of the client's response to treatment. Yet the therapist's ability to make accurate prognostic assessments, even late in therapy, has been called into question (Breslin, Sobell, Buchan, & Cunningham, 1997), especially with clients who show deterioration (Lambert & Ogles, 2004). Monitoring of treatment response remains largely at the level of intuition and clinical experience. Grove and Meehl (1996) suggest that clinician experience makes little or no difference in predictive accuracy relative to actuarial data; Dawes (1989, p. 465) concludes, "The validity of **?1?** clinical judgment and amount of clinical experience are unrelated." However, the fact remains that mental health experts often justify predictive judgments on the basis of their years of experience. This situation is particularly troubling when coupled with the research evidence that therapists are reluctant to recognize deterioration, tend to overestimate improvement rates (Norcross, 2003), and are inclined to devalue actuarial/statistical data.

Failure to use decision-making technologies is typical in psychotherapy despite the fact that research can provide lab tests on the basis of information that is at least as good as that used in other areas of health care, including medicine. For example, psychotherapy practice can be informed by research on patterns of change over the course of therapy and evidence that early response to therapy predicts final outcome. For example, Richard and Kordy (in press) noted four distinct patterns of change over the course of treatment **?2?** for bulimia, two suggesting a positive final outcome and two suggesting treatment failure. These patterns can be recognizable by the fourth week of treatment. Tang and DeRubies (1999a, 1999b), as another example, found that rapid (dramatic) improvement in depressive symptoms foretold better ultimate outcome and follow-up functioning than slow improvement. Haas, Hill, Lambert, and Morrell (2002) found the same trend in a group of clients who had a variety of disorders. Positive treatment response within the first three treatment sessions occurred in the majority of clients who improved by the end of treatment and maintained gains 6 months to 2 years after termination. There are clear and practical consequences of early poor response, and we can modify treatment before termination for these clients (Wilson, 1999).

In contrast, some patterns of treatment response noted by clinicians are inconsistent with the empirical evidence. One such pattern is that clients become worse before they become better, and that this worsening is to be expected and is a positive sign that therapy is working. This pattern is, in fact, a rare pattern in routine care, occurring in less than 10% of the cases that were studied (Canen & Lambert, 1999). The pattern was even more rare among clients who ultimately had a positive treatment response. Although infrequently studied, because therapists' ability to recognize likely treatment failure has clear implications for effective practice, understanding it is important.

In this article, we summarize our attempts to develop a lab test and statistically derived cut scores for the purpose of identifying potential treatment failures and thereby supplementing therapist judgment and decision making. We then report the results of a study in which we contrast the effectiveness of the lab test procedures with judgments made by clinicians.

Development of a Useful Lab Test

A metric expressing the degree of client disturbance was developed for the purpose of measuring treatment response on a weekly basis and at termination. The Outcome Questionnaire-45 (OQ-45) was designed to assess three aspects of the client's life: subjective discomfort/symptoms, problems in interpersonal relationships, and problems in social role performance. The items also measure personally and socially relevant characteristics that affect the individual's quality of life, attempting to quantify both positive and negative functioning. Each item is scored on a 5-point scale with the total score yielding a range of possible scores of 0 to 180; higher values indicate the endorsement of pathology. The OQ-45 was first published in 1994 and has steadily grown in usage. In fact, a 2004 survey indicated that it has become the third most frequently used measure of **?3?** treatment outcome by psychologists in clinical practice (Hatfield & Ogles, in press). **?4?**

Completion of the OQ-45 takes approximately 5 to 7 minutes, and it is typically administered before each treatment session.The OQ-45 has adequate internal consistency ($r = 0.93$) and 3-week test-retest reliability ($r = 0.84$). Concurrent validity is moderate to high ($r = 0.50–0.85$) when correlated with measures most often used to assess psychotherapy outcome in clinical trials (Lambert et al., 2004). Most importantly, the OQ-45 has been shown to be sensitive to change in clients over short periods while remaining stable in untreated individuals (Vermeersch, Lambert, & Burlingame, 2000).

Using formulas developed by Jacobson and Truax (1991), the normative data for the OQ-45 have been analyzed to provide cutoff scores for both a reliable change index **?5?** (RCI) and clinically significant change. Using normative data from community nonclients ($N = 1,353$) and clients entering treatment ($N = 1,476$), the RCI was estimated to be 14 points. Clients who change in a positive or negative direction by at least 14 points are regarded as having made reliable change. The cutoff on the OQ-45 for demarcating the point at which a person's score is more likely to be from the dysfunctional population than a functional population was estimated to be 64. This score turns out to be about a standard deviation higher than the mean of the nonclient sample ($M = 45.31$; $SD = 19.42$). When a client's score falls at 63 or below, his or her functioning is, at that point in time, considered to be more like that of nonpatients than of patients (Lambert et al., 2004). When the client improves by 14 or more points and passes below 64, the client has met criteria for making clinically significant change.

Producing cut scores for normal functioning is essential for monitoring client treatment response as it provides an individualized, quantitative, and norm-based marker for both research and clinical use. In addition, development of a method for predicting final treatment status is essential. We developed a purely statistical method and an empirically **?6?** derived method for predicting final treatment status. The data were drawn from a national database for research using the OQ-45. The treatment settings included training clinics, university counseling centers, employee assistance programs, outpatient clinics, and private practice. In most cases, clients from their respective samples filled out an OQ-45 at each treatment session. Clients who had only one treatment session or one OQ-45 administration were dropped from the data pool because no recovery curves could be generated

from only one data point. The resulting aggregate sample was 11,492 clients with two or more OQ-45 administrations.

First, the full range of scores (0–180) was divided into distinct groups on the basis of severity of disturbance. Fifty groups were formed, identified by intake score, with no fewer than 220 clients in each band, representing approximately 2% of the total sample. The resulting data were analyzed to generate a linear model for expected treatment response.

Next, the data were modeled to produce tolerance intervals, a quality control protocol often used in engineering applications. Tolerance intervals determine the probability that a given OQ-45 score at a given session would fall within a specified interval. With large data sets, the estimated upper and lower limits are equivalent to confidence intervals. Thus, the tolerance intervals allowed for the identification of OQ-45 total score values that have an established probability of falling outside the upper and lower limits of the tolerance interval. Specifically, we were able to identify the 10% of clients, at a given level of disturbance, whose rate and trajectory of progress deviated significantly from the predicted course of recovery for those entering therapy with a similar intake score. The tolerance intervals created were primarily aimed at identifying this 10% of the patient population who deviated from the recovery track. A tolerance band was also established for the 15% most poorly responding clients.

These tolerance intervals formed the core of the empirically derived warning system by providing table values and charts of predicted therapeutic gains, against which any given client at any level of disturbance at any given treatment session could be compared. If at sessions after intake the OQ-45 total score for a client was above the 10% or 15% tolerance interval, then treatment response was deemed to be deviating from expected treatment response. The therapist would receive either a red or a yellow message, depending on the degree of departure from the expected treatment response. We view the statistically derived empirical method as a lab test that can assist clinicians in evaluating the effectiveness of the current treatment.

How Well Can an Empirical Method Predict Treatment Failure?

In order to determine whether the cut scores for tolerance bands were able to identify treatment failures, 492 consecutive clients who received psychotherapy at a college counseling center were followed during the course of therapy. Their average pretreatment OQ-45 score was 72.77, with a standard deviation of 22.19. Their average number of therapy sessions was 4.44 ($SD = 3.06$).

Psychotherapists were 49 counseling center staff consisting of 22 doctoral-level students in training, including interns, and 27 licensed professionals, mainly psychologists. Therapists had a variety of theoretical orientations, and most subscribed to an integration of two or more theoretical systems. The most common single orientations were cognitive/behavioral (50%), psychodynamic/interpersonal (20%), and humanistic/existential (20%).

Clients took the OQ-45 before each therapy session. Their data were stored electronically and were not shared with the therapist or client. For the purpose of this study, treatment failure was defined as reliable worsening (deterioration), specifically a negative change of 14 or more points. In the case of clients who started therapy in the functional range (i.e., 63 or below on the OQ-45), in addition to worsening by 14 or more points, they had to leave therapy with a final score of 64 or higher, which placed them in the dysfunctional range.

The results of this analysis are presented in Table 1, which compares predicted and actual treatment failures. As can be seen in the right-hand column, titled Total Score, of the 492 clients who entered treatment 36 (7.3%) were reliably worse/deteriorated at

Table 1
Comparison of Predicted Versus Actual Treatment Failure by an Empirically Derived Classification Method

	Predicted Positive Outcome	Predicted Negative Outcome	
	N (%)	*N* (%)	Total (%)
	(Hits)	(False Negative Results)	
Actual positive outcome	273 (81.8%)	83 (18.2%)	356 (92.7%)
	(False positive results)	(Hits)	
Actual negative outcome	0 (0%)	36 (100%)	36 (7.3%)
Total number classified	273 (75.8%)	119 (24.2%)	492 (100%)
	N (%)		
Overall hit rate	409 (83.1%)		

termination. The empirical method correctly identified all 36 (100%), 86% of whom were identified by the third treatment session. The empirical method was highly effective in accurately identifying clients who went on to deteriorate.

At the same time, the empirical method misidentified 83 (18%) clients as likely to have a negative outcome, whereas they did not. The outcome of these misidentified patients (false alarms/false positive results) was further studied and contrasted with the outcome of clients who were not identified as signal-alarm cases (predicted positive outcome). Of the 83 misclassified signal-alarm cases, 18% of those patients improved or recovered at termination, and 74% showed no reliable change. In contrast, of the 373 clients whom the empirical method did not identify as signal-alarm cases, 50% recovered or improved and 50% showed no reliable change. These findings offer further support for the signal-alarm method in that they suggest that even the false alarms have a poorer outcome than cases that are not identified as likely treatment failures. That is, if an alarm is given the client has a 1/5 chance of having a positive outcome, compared to a 50/50 chance if no alarm is given.

Unlike some medical decisions in which the cost of overidentification of signal cases may result in intrusive and even health threatening interventions, the signal-alarm in psychotherapy merely alerts the therapist to the need for reconsidering the value of ongoing treatment, rather than mandating specific changes. Thus, we see the signal-alarm as supporting clinical decision making, rather than supplanting it. Because the signal-alarm alerts therapists to the possible need for action, rather than triggering a negative chain of events such as termination or referral, the current level of misidentification would seem to be tolerable.

Further analyses explored the difference between red and yellow warnings: What was the relative outcome for clients who received a red versus a yellow signal? Outcome for these clients was classified into three categories: reliably improved/recovered, no reliable change, and deteriorated. Of the 36 deteriorated cases, the empirical method's red alarm picked up 34 of the 36 clents; the yellow signal picked up the remaining 2. The red alarm is indeed a more serious indicator for deterioration, one that should generate greater concern among clinicians than the yellow signal.

Further research is needed to examine the accuracy of classifying outcome with other, more comprehensive measures and methods. The problems with undertaking such

research are the burden it places on clients to undergo repeated testing and the need to keep such research within the bounds of routine clinical practice. We argue that it is essential to employ and study methods that can actually be used in clinical practice rather than elaborate methods that cannot be easily applied in clinical settings.

To illustrate further the need for a good lab test we report now on a study that compared the accuracy of therapist predictions and the empirical method predictions of ultimate treatment failure.

How Well Can Therapists Predict Treatment Failure?

To investigate therapist accuracy in predicting negative treatment outcomes we asked a group of 48 therapists (26 trainees and 22 licensed staff) at a university outpatient clinic to predict which of their clients were likely to end treatment worse off than when they started treatment. Routine practice in this counseling center included weekly administration of the OQ-45 with provision of feedback to therapists on client progress on a weekly basis over the succeeding 4 years. The psychotherapists were therefore familiar with the **??7?** clientele and the outcome measure. The routine feedback procedure was temporarily suspended for 3 weeks (and OQ-45 scores were stored electronically and not revealed to the therapist or client) in order for therapists to fill out the prediction questionnaire that was provided for each client seen during the 3 weeks of the study.

The purpose of the questionnaire was to assess each therapist's ability to estimate client *progress* as compared to the weekly self-report measure as well as to predict client *outcome* at termination. Specific questions were as follows: (1) In your clinical judgment alone, predict this client's end of treatment outcome. This client will (choose one prediction) Recover, Improve but not recover, Make no progress in treatment, Get worse; (2) Considering this client's initial session with you, rate this client's progress as of today's session. Base your rating on your clinical judgment and clinical experience alone. Today this client is (choose one rating) Recovered and ready for termination; Improving as expected, but in need of continued treatment; Making no progress or poor progress; Getting worse.

Therapists were informed both verbally and in written form before the study that deterioration rates in the clinic had remained relatively constant at 8% over the preceding years, and that our primary interest in administering the questionnaires was to learn whether therapists could indeed predict that important percentage of clients who worsen during psychotherapy, as indicated by the OQ-45 posttreatment score. Clients in the study were followed with respect to their weekly OQ-45 self-report throughout the course of therapy.

Predictions were made for 618 clients seen during the 3-week period. Clients were seen for individual therapy, couples therapy, and crisis (walk-in) intervention. After removing those clients who did not have OQ-45 session data that allowed treatment progress and outcome to be estimated ($n = 18$), or who did not take the measure ($n = 15$), or who had only an intake session ($n = 35$), data for 550 clients could be analyzed. It should be noted that some clients in the study were being seen for their initial visit and some had already had several sessions before the prediction questionnaire administration. In addition, some clients had more than one session during the 3-week period and so therapists made more than one rating of their progress or likely outcome. Therapists made between one and four predictions for clients, resulting in a total of 944 predictions for the 550 clients. Of the 265 clients who had more than one prediction, if the client was predicted to deteriorate by the end of treatment at any measurement point, we considered it a prediction of a negative outcome. The questionnaires resulted in a cross-sectional sample

of predictions across different therapy sessions. In all, data were obtained for 218 clients who completed *intake sessions and first sessions* and 332 clients who completed two or more sessions.

Results for predicted outcome compared to actual outcome are presented in Figure 1. *Deterioration* was defined as reliable negative change, *positive outcome* was defined as clinically significant or reliable improvement, and *no change* or *no progress* was defined as any score that showed neither reliable improvement nor reliable deterioration.

The results in Figure 1 illustrate that therapists rarely predicted deterioration. Only 3 (0.01%) of 550 clients were predicted to deteriorate, and only 1 of those predicted to deteriorate had, in actuality, deteriorated at the end of therapy. Actual outcome data indicated that 40 clients (7.3%) deteriorated by the end of therapy. We interpret these results as indicating that therapists tend to overpredict improvement and fail to recognize clients who worsen during therapy. This occurred even when therapists knew the baseline deterioration rate and were familiar with the outcome measure that was used as the criterion for categorization of outcome. In fact, the failure of the therapist to predict negative outcome precluded us from analyzing some questions of interest, such as the difference between experienced and inexperienced psychotherapists. The simple finding was that therapists in this study simply did not anticipate negative treatment outcomes.

A second question of interest was, Is there agreement between therapist-judged lack of progress and the empirical-method-derived lack of progress? The subsample of interest in this analysis were those clients whom the therapist had seen on multiple occasions and therefore whose response to therapy they had an opportunity to assess. Because the empirical method of predicting treatment failure relies on information about treatment response (client progress), we can also compare the empirical (lab test) method to therapist judgments. Comparisons among therapist-predicted outcome, lab test-predicted outcome, and actual outcome are presented in Figure 2.

To compare therapist predictions of outcome with lab test/actuarial predictions collected at the same time, we analyzed data from only 332 clients (recall that 218 of the 550 clients were being seen at intake or for the first time by the therapist, or did not take an OQ-45 at the session when therapists made predictions, making progress prediction

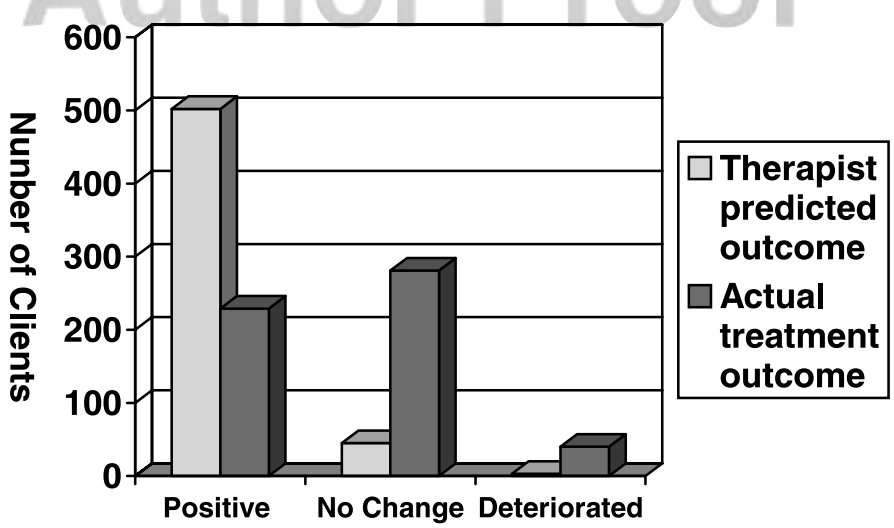

Figure 1. Therapist predicted treatment success compared to actual treatment outcomes after psychotherapy.

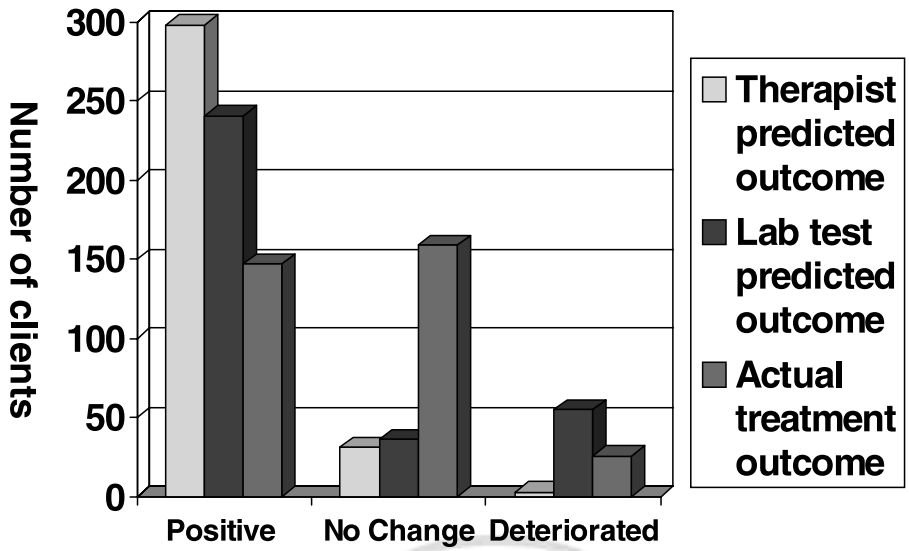

Figure 2. Comparison of therapist-predicted, lab-test-predicted, and actual client outcomes after psychotherapy.

impossible for these 218 clients). In the case of multiple OQ-45 and therapist predictions for a single client, the client could be rated as progressing but not recovered, not changing, or getting worse. These ratings were independent of predictions of final outcome. For example, a therapist could see the client as not progressing at this time but still likely to achieve a successful outcome. Of the 332 clients who had both therapist and OQ-45 progress ratings, 26 clients (7.8%) actually deteriorated, a figure consistent with past findings.

Although therapists judged that 16 clients (5%) had worsened at the time of a particular session, they predicted outcome for all but 1 of the 16 either to be positive or, at worst, to show no reliable change by termination. Across the 332 clients, therapists predicted only 3 clients would deteriorate at final outcome, of whom 1 was correctly predicted to be worsening at the time of prediction and was also correctly predicted to deteriorate at the end of treatment.

When multiple predictions and multiple OQ-45 scores were available for clients seen more than once during the study dates, the most conservative (i.e., negative) prediction was used for both therapist-judged and empirically judged poor treatment response. Results indicate that, of the 26 clients who deteriorated, during the 3-week window, the lab test identified 55 clients as *potential* treatment failures, 20 (77%) of whom did indeed deteriorate by the end of treatment. In contrast, therapists identified 5 of the 26 deteriorated clients as having worsened during therapy (at the time of the prediction) but correctly identified only 1 client as a potential treatment failure (0.04%).

These results reinforce the notion that therapists need independent data to alert them that treatment is not having its intended effects and that deterioration may be forthcoming. These results are consistent with past research on clinical versus actuarial predictions (Dawes, 1989; Grove & Meehl, 1996) and support the conclusion that formal methods of monitoring client treatment response may be of practical value.

Practice Conclusions

Positive outcomes can be expected for most clients who enter psychotherapy, and clinicians are appropriately optimistic about the effects of their services. Despite clinician con-

fidence in their ability to help clients, many clients do not benefit and a small portion even deteriorate before they leave treatment. It appears that clinicians are unwilling or unable to identify these clients, despite knowing the baseline rates for failure. Whereas other health care professionals rely heavily on lab test results for diagnosis and monitoring of treatment response, psychotherapists continue to practice without reliance on them. Formally tracking treatment response with the use of standardized measures and markers for likely negative response increases the likelihood that clinicians can take timely steps to reduce treatment failure. We now have the ability to use effective lab tests to aid clinical practice. The nagging question is whether clinicians will use them in routine practice. We have developed elegant methods for accurately predicting client treatment response, and failure to respond in particular. The degree to which such information, when made accessible to practitioners, improves client outcome is a topic highlighted in other articles in this issue.

Select References/Recommended Readings

Breslin, F., Sobell, L.C., Buchan, G., & Cunningham, J. (1997). Toward a stepped-care approach to treating problem drinkers: The predictive validity of within-treatment variables and therapist prognostic ratings. Addiction, 92(11), 1479–1489.

Dawes, R.M. (1989). Experience and validity of clinical judgment: The illusory correlation. Behavioral Sciences and the Law, 7(4), 457–467.

Finch, A.E., Lambert, M.J., & Schaalje, B.G. (2001). Psychotherapy quality control: The statistical generation of expected recovery curves for integration into an early warning system. Clinical Psychology and Psychotherapy, 8, 231–242.

Grove, W.M., & Meehl, P.E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. Psychology, Public Policy, and Law, 2(2), 293–323.

Haas, E., Hill, R., Lambert, M.J., & Morrell, B. (2002). Do early responders to psychotherapy maintain treatment gains? Journal of Clinical Psychology, 58, 1157–1172.

Hatfield, D., & Ogles, B.M. (in press). The use of outcome measures by psychologists in clinical practice. Professional Psychology: Research & Practice. **?8?**

Jacobson, N.S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology, 59, 12–19.

Lambert, M.J., Morton, J.J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R.C., et al. (2004). Administration and scoring manual for the Outcome Questionnaire (OQ-45.2). Orem, UT: American Professional Credentialing Services.

Lunnen, K.M., & Ogles, B.M. (1998). A multiperspective, multivariable evaluation of reliable change. Journal of Consulting and Clinical Psychology, 66(2), 400–410.

Mohr, D.C. (1995). Negative outcome in psychotherapy. Clinical Psychology, 2, 1–27.

Norcross, J.C. (2003). Empirically supported therapy relationships. In J.C. Norcross (Ed.), Psychotherapy relationships that work (pp. 3–16). New York: Oxford University Press.

Richard, M., & Kordy H. (in press). Early treatment response: Conceptualization, predictive validity and application in quality management. **?8?**

Tang, T.Z., & DeRubies, R.J. (1999a). Reconsidering rapid early response in cognitive behavioral therapy for depression. Clinical Psychology: Science and Practice, 6(3), 283–288.

Tang, T.Z., & DeRubies, R.J. (1999b). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. Journal of Consulting and Clinical Psychology, 67(6), 894–904.

Vermeersch, D.A., Lambert, M.J., & Burlingame, G.M. (2000). Outcome Questionnaire-45: Item sensitivity to change. Journal of Personality Assessment, 74(2), 242–261.

Wilson, G.T. (1999). Rapid response to cognitive behavior therapy. Clinical Psychology: Science and Practice, 6, 289–292.